Learning vocabulary with the support of sustained exposure to captioned video: do proficiency and aptitude make a difference?

Maria del Mar Suárez & Ferran Gesa

Published online: 24 May 2019.

Submit your article to this journal

View Crossmark data

Routledge
Taylor & Francis Group

Check for updates

Learning vocabulary with the support of sustained exposure to captioned video: do proficiency and aptitude make a difference?

Maria del Mar Suárez [a] and Ferran Gesa [b]

aDepartment of Language and Literature Education, Universitat de Barcelona, Barcelona, Spain; bDepartment of Modern Languages and Literatures and English Studies, Universitat de Barcelona, Barcelona, Spain

ABSTRACT

Video viewing can be a valuable resource to expose students to large quantities of input so they can improve their vocabulary and content comprehension. Most studies so far have used short clips and have not explored in much detail the effects of individual differences (IDs) such as aptitude, listening skills and vocabulary size. This paper aims to address this gap by exposing 57 Grade-10 EFL learners and 60 university students to captioned video. On a weekly basis over an academic term, all learners were pre-taught a set of target words (TWs); half of them (the experimental group) were additionally shown captioned episodes from a TV series containing the TWs. All learners were pre- and post-tested on the TW forms and meanings. Results revealed significant differences between experimental and control groups in the learning of TWs in the high school population, but not among university participants. A main effect for proficiency was observed on the learning scores for both TW forms and meanings. However, language aptitude was only a significant factor for TW meanings. Results are discussed regarding how video viewing and these IDs mediate vocabulary learning.

Introduction

Vocabulary is undoubtedly essential for language use (Nation and Waring 1997). However, the number of words taught - and hopefully learned - in classroom settings is said to be insufficient in order to really advance in foreign language (FL) learning (Malone 2018). Thus, students need to look for alternative sources of input in order to progress in their FL development. Among these other sources of input, recent research has begun to examine the potential for L2 vocabulary development of television viewing (Peters and Webb 2018; Rodgers 2013), particularly with captions (Montero Perez, Van den Noortgate, and Desmet 2013; Montero Perez et al. 2014; Sydorenko 2010). Most studies so far have demonstrated a positive effect of FL television for vocabulary learning (e.g. Winke, Gass, and Sydorenko 2010); however, captions have not always been proven to be beneficial for this area of language learning (Rodgers 2013), potentially leading to cognitive overload at beginner levels which, in turn, hinders language acquisition (Mayer, Lee, and Peebles 2014).

Vocabulary size, or 'how many words a learner knows' (Coxhead, Nation, and Sim 2015: 121), has been shown to have a positive effect on areas of language learning such as reading (Laufer and Ravenhorst-Kalovski 2010; Nation 2006; Webb and Chang 2015) or listening and writing (Staehr 2008). This positive relationship has also been observed in vocabulary learning through (captioned) video viewing (Montero Perez, Peters, and Desmet 2018; Peters, Heynen, and Puimège 2016; Peters and Webb 2018), in what might be seen as an instance of the Matthew effect,[1] or the 'rich-get-richer'

---

principle, according to which higher ability students tend to make greater language gains (Penno, Wilkinson, and Moore 2002; Stanovich 2009). Similarly, language aptitude has been linked to greater vocabulary knowledge (e.g. Dahlen and Caldwell-Harris 2013; Kormos and Sáfár 2008). So far, SLA research has looked into the connections between vocabulary learning, language aptitude and proficiency in many ways, but to our knowledge, not in a classroom in the context of learning vocabulary from viewing captioned TV series.

The aim of this study is thus twofold: first, it aims to examine whether captioned video viewing, supporting teacher-led instruction[2] can lead to higher vocabulary gains than teacher-led instruction only; second, it explores the influence on vocabulary learning supported by captioned video of L2 proficiency (operationalised as vocabulary size and listening skills) and language aptitude.

Literature review

Vocabulary learning through (captioned) video

In FL contexts (i.e. learning the language in a classroom context in a country in which it is not spoken by the community), where exposure to the target language (TL) is typically limited to a few hours per week (Muñoz 2008), learners need to find additional sources of input and new opportunities for exposure to the TL. Among these, viewing (captioned) television is one of the most prevalent sources of input for FL learning (Lindgren and Muñoz 2013). Given the widespread use of audiovisual media and the growing popularity of audiovisual materials in language teaching, research has focused on investigating the extent to which foreign language acquisition can be boosted by multi-modal input (Ghia 2012; Mitterer, and McQueen 2009; Vanderplank 2010, 2016).[3]

Most research so far has been directed to the role that captions or L2 subtitles[4] play in learning FL vocabulary, almost neglecting the question of whether video viewing alone, with no written support, can promote increased learning. In this respect, only two studies that we are aware of have compared vocabulary learning with, and without, video viewing. Rodgers (2013) investigated the incidental vocabulary acquisition of a set of 60 TWs in Japanese (pre-)intermediate EFL undergraduates who watched ten episodes of an American television series. Results showed a small but significant impact for watching English-language television, indicating that encountering the words in context and with visual support promoted incidental vocabulary acquisition. More recently, Peters and Webb (2018) looked at vocabulary acquisition through video viewing in Flemish intermediate EFL learners. Learners in the experimental condition watched a one-hour long documentary while the control group just took the vocabulary tests. All participants were tested on the meaning recognition and recall of 64 TWs one week before the experiment and immediately after it. Statistical analyses revealed significant differences between the two conditions, with video viewing explaining 21% of the variance on the meaning recall test and 8% on the meaning recognition test. However, the authors call for further research because their results were not felt to be generalisable to other populations due to the strong presence of English in Flanders, where television and movies are not dubbed (European Commission 2011) and people are exposed to original version audiovisual material from an early age – not typically the case in other areas of the world.

Whether captions, or any other kind of subtitling, facilitate vocabulary learning has also been a focus of investigation over the past years. In this respect, a state-of-the-art review (Vanderplank 2016) and a meta-analysis (Montero Perez, Van den Noortgate, and Desmet 2013) have confirmed the benefits, but also the limitations, of captioning for FL learning. Captions appear to be beneficial for specific areas of language development, in particular, vocabulary acquisition, listening comprehension, learning strategies, speech segmentation and literacy development. Their impact on two of these areas (vocabulary acquisition and listening comprehension) was systematically reviewed in Montero Perez, Van den Noortgate, and Desmet's (2013) meta-analysis. With respect to vocabulary learning, the authors selected ten research studies (all of which compared captioned and uncaptioned video viewing) and found a large effect size for captioning on the vocabulary tests

administered, supporting the idea that 'learners who were exposed to captioned video significantly outperformed learners in the control group (*p*< .001)' (730). Captioning proved to be equally effective on recognition and recall tests. However, these results should be interpreted with caution as, ideally, a larger number of research studies should have been included in the sample, in particular, on vocabulary recognition. The authors concluded that captioning was beneficial for all levels of proficiency, provided that the input matched learners' actual interlanguage level.

More recently, research has found that watching video with different types of captioning (full, keyword, full with highlighted keywords or glossed) was conducive to higher scores on FL vocabulary tests than watching uncaptioned videos. This was especially true for form recognition and clip association tests, which assessed whether participants associated words with the corresponding clip. Progress was harder to observe on more demanding tests such as meaning recall (Montero Perez et al. 2014; Montero Perez, Peters, and Desmet 2018). Some studies (e.g. Sydorenko 2010) have nevertheless found that captions are indeed beneficial for the learning of TW meanings, rather than simply promoting higher scores on form recognition tests. Captions have also been shown to be beneficial for learning both written and aural forms of words, as indicated by Winke, Gass, and Sydorenko's (2010) results analysing the behaviour of second-year university learners of Spanish watching three short videos, either with, or without, captions. However, other studies have found uncaptioned video viewing to be equally effective, resulting in the same amount of vocabulary learning as captioned viewing (Galimberti 2016; Rodgers 2013; Yuksel and Tanriverdi 2009). In one case (Hsu 2013), viewing uncaptioned video even led to greater use of advanced vocabulary in participants' writing than viewing captioned videos.

The effect of vocabulary size and listening skills on vocabulary learning from (captioned) video

Research has been able to identify a positive significant relationship between vocabulary size (VS) and different aspects of lexical knowledge: form and meaning recognition (Montero Perez et al. 2014; Peters, Heynen, and Puimège 2016), form recall (Peters, Heynen, and Puimège 2016) and meaning recall (Montero Perez et al. 2014; Montero Perez, Peters, and Desmet 2015; Peters, Heynen, and Puimège 2016). It is thus reasonable to conclude that 'the effect of this parameter [VS] seems to be quite robust, as the effect of vocabulary size was about the same in all tests, viz. the odds of a correct response increased 2–5% for one word known more in the vocabulary size test' (Peters, Heynen, and Puimège 2016: 145). These results were confirmed by Peters and Webb (2018), who found that ten more words known on the VS test increased by 32% the odds of getting a correct response on meaning recall and form and meaning recognition tests. Likewise, a positive relationship has also been observed between prior vocabulary knowledge and scores on immediate vocabulary tests tapping into colloquial language use (Frumuselu 2015). A higher proficiency level was also found to lead to more vocabulary uptake in studies comparing two groups of students at different ages (taken as a proxy for different proficiency levels) in the same instructional context and following the same experimental design (Alexiou 2015; Aurstad 2013; Kvitnes 2013).

No studies to date that we are aware of explore the effects of listening skills on vocabulary learning through video viewing. Previous research examining the relationship between vocabulary and listening has shown that vocabulary size is significantly, although modestly, correlated with listening comprehension (Noreillie et al. 2018; Staehr 2009). This also holds true as far as the relationship between vocabulary depth and listening comprehension is concerned (Staehr 2009). However, when comparing listening with other language skills, it correlates with vocabulary to a lesser extent than reading or writing (Staehr 2008) or even speaking (Miralpeix and Muñoz 2018). In this latter study, for example, the authors found that listening was the least correlated skill with vocabulary size, explaining 15.6% of the variance in the vocabulary size scores. Listening activities are nevertheless often suggested as a good source of input for vocabulary learning (Elley 1989; Maneshi 2017; van Zeeland and Schmitt 2013; Vidal 2003) and retention (Vidal 2011).

Language aptitude and vocabulary learning

Language aptitude has been defined as a set of cognitive abilities that are 'predictive of how well, relative to other individuals, an individual can learn a foreign language in a given amount of time and under given conditions' (Carroll and Sapon 2002: 23). Although it has consistently been considered a multi-componential construct, the actual elements that contribute to the aptitude construct are debatable. The traditional conceptualisation of aptitude (Carroll 1981) considers that it consists of phonemic coding, inductive language learning, grammatical sensitivity and rote learning. Apart from inductive learning, these components are tested in various subtests comprising the most widely used aptitude test to date, the Modern Language Aptitude Test (Carroll and Sapon 1959), and revisited in a newer, freely available test designed for research, that is, the LLAMA test (Meara 2005a). This test involves the following sub-tests, each of which is designed to measure different learning abilities:

- LLAMA B – vocabulary learning: test-takers are expected to learn 20 words associated with a target image in two minutes. In the testing phase, test-takers are instructed to match the name of an object with its drawing.
- LLAMA D – sound-pattern recognition: after listening to a string of ten sound sequences, test-takers are expected to discriminate between previously heard or unheard sound strings.
- LLAMA E – sound-symbol association: test-takers learn the sound of 24 written syllables for two minutes and then have to associate them with two-syllable written combinations.
- LLAMA F – grammatical inferencing: test-takers have to infer the rules of an unknown written language from the principles underlying a set of pictures representing them.

Other recent conceptualisations of aptitude have attempted to include other constructs, such as working memory (Miyake and Friedman 1998) and phonemic short-term memory (Bolibaugh and Foster 2013), and try to take into consideration such factors as the influence of implicit and explicit learning processes (Kaufman et al. 2010).

The overall score from an aptitude test such as the MLAT gives a general indicator of aptitude. However, different sub-components or 'aptitudes' (Carroll 1993: 675) may well be specifically helpful in mastering different aspects of language knowledge and skills, and thus, overall aptitude scores may not necessarily be strongly associated with a specific skill such as vocabulary learning. In a recent meta-analysis, Li (2016: 801) reported that 'aptitude measured using full-length [aptitude] tests was a strong predictor of general L2 proficiency, but it had low predictive validity for vocabulary learning and L2 writing'. In any case, aptitude research has typically tended to focus on L2 morpho-syntactic learning (Saito 2017) and only a few relevant studies have analysed the association of aptitude with vocabulary learning: these include Granena and Long (2013) on the role of aptitude in lexis and collocations learning in adults with age of onset 16–29; Dahlen and Caldwell-Harris (2013) on word recall and recognition; and Saito (2017) on lexical frequency and richness.[5] Aptitude tests are designed to predict rate of language learning to account, at least in part, for why in similar conditions, some learners will advance quicker than others. However, research shows that while aptitude scores do appear to be highly predictive of rate of progress at early language learning stages (Doughty 2019), the role of aptitude seems to diminish as L2 proficiency improves, presumably because more proficient learners can resort to other cognitive skills and strategies to advance their language learning (Serafini and Sanz 2016; Winke 2013).

Language aptitude and implicit and explicit learning

The constructs of explicit and implicit knowledge have been widely researched in SLA. Explicit language knowledge is derived from an active and conscious learning process whereas implicit language knowledge develops without necessary awareness (Andringa and Rebuschat 2015; Williams 2009). There has been significant debate as to whether aptitude is relevant only in the case of explicit

learning, or whether it plays a role in implicit learning too (see Skehan 2012). It is indeed plausible that different 'aptitudes' are involved in the two different learning processes. This was in fact highlighted by Granena's (2013) study, where principal components analyses of the tests comprising LLAMA identified an explicit learning component, based on the LLAMA B, E and F subtests, and an implicit learning component, based on LLAMA D. Nevertheless, a recent meta-analysis (Li 2015) concluded that aptitude seemed more likely to be involved in explicit contexts, though Hwu and Sun (2012) found that, in high aptitude learners, aptitude did not seem to play a differential role in explicit vs. implicit contexts. In the specific case of L2 vocabulary learning in FL contexts, it seems likely that explicit learning is involved to a great extent, despite the fact that we acquire much of our L1 vocabulary incidentally. This is no doubt due to the limitations of the FL context, where there is typically insufficient input for sustained incidental vocabulary learning to occur (Webb and Nation 2017). It is reasonable to assume, therefore, that aptitude may well be a factor influencing effective L2 vocabulary learning.

The present study

Aims and research questions

Research so far has shown that captioned video viewing is generally beneficial for language learners. However, 'there are few substantial studies that have involved watching captioned videos or films over a period of a few weeks in order to assess changes of behaviour and gains over time' (Vanderplank 2016: 117; see Bravo (2008) or Rodgers (2013) for exceptions). One of the aims of this study is thus to investigate the effect on vocabulary learning of sustained exposure to captioned videos in the context of formal teaching. Previous research has also mainly focused on university learners (e.g. Etemadi 2012; Winke, Gass, and Sydorenko 2013) so this study also includes a younger lower-proficiency group of participants (specifically Grade 10 students, aged 15–16). L2 proficiency has been shown to impact directly on vocabulary learning, but the role of FL aptitude has not been investigated in detail. This study, therefore, also sets out to explore the impact of both proficiency (as measured by participants' listening skills and vocabulary size) and FL aptitude (in its traditional conceptualisation as measured by the LLAMA test) on vocabulary learning through captioned video viewing as part of formal classroom-based language learning. Our research questions are as follows:

In high school and university EFL learners:

(1) does sustained exposure to captioned episodes from a TV series in the context of formal language instruction lead to significant gains in vocabulary learning, compared to receiving formal language instruction only?
(2) to what extent do listening skills, vocabulary size and language aptitude mediate any gains in vocabulary learning from viewing captioned episodes from a TV series in the context of formal language instruction?

Participants

Participants were recruited from two different educational contexts: high school, representing an under-researched population in this area (Plonsky 2015), and university, for comparison purposes. Only those students who completed the full range of experimental tasks were included in the analysis for this study, resulting in a total of 117 participants: 57 learners (28 males and 29 females), aged 15–16, enrolled in their last year of compulsory secondary education, and 60 first-year university students (21 males, 39 females) aged 18–26. Intact classes were respected in assigning participants to the experimental (EG) and control group (CG) (see Table 1).

All learners were Catalan-Spanish bilinguals studying English in public institutions in Barcelona and its metropolitan area. The high school students had received at least 1,100 h of formal instruction in the TL and were expected to be at a low-intermediate level. According to their scores on the Oxford

Table 1. Distribution of participants.

Level	EG (*n*)	CG (*n*)	Total (N)
High school	30	27	57
University	37	23	60

Placement Test (OPT), their average level was B2 on the Common European Framework of Reference for Language (CEFR), as shown by the correspondence between OPT scores and the CEFR (Allan 2004). The university students were enrolled in their first year of Media Studies at the University of Barcelona. They had received at least 1,300 h of formal instruction and so were expected to be at an upper-intermediate or low-advanced level, although within-group variability varied from A2 to C1 according to the CEFR.

Instruments

Video episodes
Eight episodes from the TV series *I Love Lucy* (Oppenheimer and Arnaz 1951) were selected. The episodes were presented to the EG in English and were accompanied by English captions. The episodes ranged between 22 m 15 s and 24 m 57 s, the average length being 24 m 30 s. The total viewing time was 3 h 15 m 59 s, excluding opening and ending themes. This TV series was chosen as it had been used with university students (Cokely and Muñoz forthcoming), who showed a positive attitude towards it. An analysis of the lexical profile of the eight episodes using VocabProfile v2.0 (Cobb ongoing) and the BNC/COCA frequency lists (Nation 2012) revealed that 95% coverage was reached at the 2,000-word level, while knowledge of the first five thousand words was needed to reach 98% coverage.

Vocabulary tests
From each of the eight episodes, five TWs were selected, making a total of forty TWs over the term. Words unlikely to be known by the participants were selected, and care was taken to avoid cognates as much as possible. Only words appearing at least twice in the specific episode were selected: see Appendix for further information on the TWs. Pre- and post-tests were then designed to measure (1) participants' productive knowledge of the TW orthographic forms in a spelling test (Webb 2007) and (2) their meaning recall through L2 to L1 translation. Our study used two tests measuring different lexical aspects (form and meaning) in order to ensure that possible vocabulary learning gains were not underestimated (Nation and Webb 2011). Spelling (or knowledge of the TW form) was tested for two reasons: participants in both groups were exposed to the English word forms through the instructional activities, and students in the EGs also saw the English written forms in the video captions (see Procedure below).

The pre-test and post-test comprised an audio recording by a native speaker, where each TW was read out twice. The presentation of items was randomised to lessen memory effects, and primacy and recency effects. At both pre- and post-test, participants were asked first to write down the English word form (spelling test). To measure meaning recall, participants were asked to provide a Spanish or Catalan translation or definition.

Formal instructional materials
Five TWs were explicitly taught in each of the eight sessions through two kinds of formal instructional materials: a vocabulary pre-task at the beginning of the session and a vocabulary post-task after watching the episode (EG) or at the end of class (CG) (see Procedure below).

The pre-tasks engaged all participants in a learning process that was intentional and explicit (Schmitt 2008). They followed a focus-on-forms approach (Laufer 2006) with different practice formats such as matching exercises, fill-in-the-blanks and crosswords; they resembled some of the

vocabulary exercises found in coursebooks. The TWs were presented in a context different from the one where they appeared in the TV series, even though care was taken to ensure that, in case of polysemous words, the meanings were the same as those presented in the TV series.

In the post-task, learners were required to listen to an audio recording of the TWs, and then write down the English word forms. They were also asked to select the best Spanish translation out of six options in a multiple-choice task, as presented below for the TW 'to sneeze'. The options appeared in randomised order so as to avoid a predictable pattern. Although it was assumed that the TWs should have been known after students had engaged with the pre-task, the 'I don't know' option was added to check whether this was actually true or not and to minimise guessing. Students were explicitly told to select it if they were not completely sure about the answer (Zhang 2013).

(1) The key [e.g. *estornudar* (to sneeze)].
(2) A semantically related distractor from the same part of speech as the TW [e.g. *resfriarse* (to catch a cold)].
(3) A distractor from the same part of speech as the TW and phonologically similar to the English form of the item [e.g. *congelar* (to freeze)].
(4) A distractor from the same frequency band and different part of speech as the TW [e.g. *superstición* (superstition)].
(5) A hapax, an item which appeared in the episode although it was not selected as a TW [e.g. *inscribir* (to register)].
(6) An 'I don't know' option [*No lo sé* (I don't know)].

Proficiency tests: listening skills and vocabulary size

All participants completed two proficiency tests: the listening part of the Oxford Placement Test (Allan 2004), containing 100 items, and the X_Lex v2.05 (Meara 2005b) and Y_Lex v2.05 (Meara and Miralpeix 2006) vocabulary size tests. In the OPT, participants are asked to listen to 100 original sentences presented by native speakers and decide which of the two options given corresponds to the words uttered in the recording. Both options are semantically and grammatically plausible, so test takers can only rely on their listening abilities to choose the most appropriate answer.

The vocabulary size tests, considered to be breadth tests, measure how many words participants know in English. X_Lex v2.05 analyses the vocabulary included in the first 5,000 words whereas Y_Lex v2.05 taps into vocabulary included in the 6,000–10,000 word range. The learners' task consists in deciding whether the word appearing on the screen is known or unknown to them. However, the tests contain a number of invented yet plausible English words; if learners claim to know one of these pseudo-words, their scores are adjusted downwards. In order to be as accurate as possible, VS scores by learners who had claimed to know six or more pseudo-words were excluded from the analysis based on Miralpeix (2012), which resulted in deleting data from 21.37% of the initial cohort of participants.

Aptitude test

The LLAMA test (Meara 2005a; see above for detail) was chosen to determine the participants' language aptitude as it is a free computer-based test widely used in SLA research and it is considered 'robust and not subject to significant external factors or individual variables that would influence their results' (Rogers et al. 2017: 56).

Procedure

The research was conducted over an academic term, during one class session a week over 11 weeks. One week before the beginning of the pedagogical intervention, participants completed the vocabulary size tests and the language aptitude test. In the second session, learners completed the listening part of the OPT and the vocabulary pre-test. This was followed by the eight intervention sessions.

Table 2. Design of the pedagogical intervention.

		Session 1 Proficiency and aptitude tests	Session 2 Vocabulary pre-test	Sessions 3–10			Session 11 Vocabulary post-test
				Vocabulary pre-task	Captioned episode viewing	Vocabulary post-task	
Group	EG	✓	✓	✓	✓	✓	✓
	CG	✓	✓	✓	×	✓	✓

As one of the aims of the study was to investigate whether captioned video viewing could enhance formal classroom-based vocabulary learning, the five TWs relevant to each week's video episode were pre-taught in class to both the control and the experimental groups. The vocabulary exercises were completed either individually or in small groups. The key to these exercises was provided orally by the teacher immediately afterwards, with the class invited to raise any doubts or questions regarding the TWs. The pre-task lasted around ten minutes. The EG then watched one captioned episode of the TV series while the CG completed other activities unrelated to the TWs for about twenty-five minutes. At the end of each session, all participants completed the post-task targeting the TWs to ensure that they had further opportunities to learn them. The completion of the post-task lasted around ten minutes and no feedback was provided in class.

One week after the eight intervention sessions, the vocabulary post-test was administered in order to be able to compute for lexical gains. The whole treatment design was integrated within the course curriculum but did not count in the participants' assessment grades. Students were at all times encouraged to answer as well as possible, and they were allocated course credits for participating in the study regardless of their scores in the tests and tasks. The design of the study is detailed in the following table (see Table 2).

Scoring

Vocabulary tests

The vocabulary pre- and post-tests were scored dichotomously, with 0 points for an incorrect answer and 1 for a correct response. For a TW form to be considered correct, no orthographic mistakes were allowed. As Webb (2007: 55) argues in relation to his intermediate level participants:

> This was because the learners were given the phonological forms of the target words as a cue to recall. Since the participants were at the intermediate level and were likely to have learned most if not all of the rules of spelling, phonological cues would be enough to at least lead them to write a close approximation of the target words. If responses with minor spelling mistakes were marked as correct, then it could not be determined whether it was due to repetition -an encounter with the target words in the tasks- or the phonological prompt.

In this respect, words like '*furnace*' or '*to hobnob*' were marked as incorrect if written '*furnac*' or '*to hopnob*', respectively. Regarding TW meanings, the criteria adopted were slightly more lenient and definitions, synonyms or translations were accepted. However, in case of polysemous words, only the meaning shown in the TV series and focused on in the vocabulary instructional materials was accepted as a correct response.[6] Less strict criteria, taking into consideration partial knowledge of TWs, were also applied but significant differences were not found between these two scoring procedures (Gesa forthcoming). For this reason, only the stricter scores are reported in the paper.

Absolute vocabulary gains (post-test minus pre-test scores) for form and meaning were calculated. To obtain a more fine-grained measure, items were then classified into two categories: learned (TW forms or meanings not known on the pre-test but known on the post-test) and known (TW forms or meanings known on both the pre- and post-test), so as to calculate relative gains. These control for the number of TWs at the item level that students already knew on the pre-test (Horst, Cobb, and Meara 1998; Peters and Webb 2018; Rodgers 2013):

$$\text{Relative gains} = \frac{N \text{ of forms or meanings learned}}{N \text{ of items} - N \text{ of forms or meanings known}} \times 100$$

Formal instructional materials

The results from the vocabulary pre- and post-tasks for each session will not be reported in this paper since they were included to give participants the opportunity to learn the TWs explicitly and do not provide a measure of vocabulary acquisition.

Aptitude tests

The maximum score for the LLAMA B, E, and F sub-tests is 100 and 75 for LLAMA D. For the sake of comprehensibility, the scores of LLAMA were transformed into percentages, as were absolute and relative gains.

Analysis

TW forms and meanings were analysed as separate constructs since they tap into different aspects of lexical knowledge (Nation 2013). The data for the high school and the university groups were analysed separately. Descriptive results for the two proficiency measures (see Table 3) showed that, irrespective of level (high school or university), the listening skills of all learners as measured by the OPT fell within the CEFR B2 level. On the vocabulary size test, the university group (with a mean score of 3,088 words on the X_Lex test) were classed at B1 level, while the high school group (2,472 words on the X_Lex test) were at the A2 level (Meara and Milton 2003; Milton 2010; Milton and Alexiou 2009).[7]

To determine the comparability of the high school and university groups, a Mann–Whitney U test was run on the listening test scores and an independent samples t-test on the VS scores, according to the normality of the data following the Kolmogorov–Smirnov test. No significant differences were found on the listening test ($U= 1,851$, $p=.441$) (university – $Median= 69.50$ vs. high school – $Median= 67.50$), but there were significant differences on the vocabulary size test ($t(105.79)= -4.329$, $p= .000$). Regarding the scores on the LLAMA tests (see Table 4 for the descriptive results), there were no significant differences between the high school and university groups, as shown by another independent-samples t-test ($t(110)=-.039$, $p= .969$). As differences arose in one of the proficiency measures administered (i.e. vocabulary size), our decision to analyse the two age groups separately was reaffirmed.

To answer the first research question – whether the experimental treatment (viewing captioned episodes from a TV series) had an effect on the learning of the target vocabulary – two Generalised

Table 3. Descriptive results of proficiency tests.

| Level | Group | X_Lex / Y_Lex | | | | Listening OPT | | | |
		Mean	SD	Min.	Max.	Mean	SD	Min.	Max.
High school	EG ($n = 30$)	3,455	922	2,100	5,600	68.37	7.62	56	92
	CG ($n = 27$)	3,496	1,009	1,350	5,750	69.07	6.55	51	83
	All ($n = 57$)	3,475	956	1,350	5,750	68.70	7.08	51	92
University	EG ($n = 37$)	4,528	1,343	1,650	7,150	69.19	6.96	57	82
	CG ($n = 23$)	4,233	1,417	2,000	7,250	70.35	8.21	56	83
	All ($n = 60$)	4,415	1,367	1,650	7,250	69.63	7.42	56	83

Table 4. LLAMA descriptive results in percentages.

| Level | Group | LLAMA B | | LLAMA D | | LLAMA E | | LLAMA F | | LLAMA Total | |
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
High school	EG ($n = 30$)	61.07	24.55	35.48	19.96	76.07	23.15	49.64	25.31	55.57	14.74
	CG ($n = 27$)	54.79	19.25	33.61	18.18	74.58	27.92	49.79	20.24	53.19	12.99
	All ($n = 57$)	58.17	22.27	34.62	18.99	75.38	25.24	49.71	22.89	54.47	13.88
University	EG ($n = 37$)	54.86	21.62	39.82	23.86	70.81	27.53	49.05	23.45	53.64	16.52
	CG ($n = 23$)	51.74	18.19	33.91	17.97	79.13	20.65	59.57	22.25	56.09	10.49
	All ($n = 60$)	53.67	20.27	37.56	21.82	74	25.26	53.08	23.38	54.58	14.46

Linear Models (GLZs) with gamma distribution were run, one testing relative gains in knowledge of TW form and the other, relative gains in knowledge of TW meaning. Level (high school vs. university) and condition (EG vs. CG) were entered as fixed effects and their simple contrasts were subsequently inspected. Data from all participants were included in the analysis ($N = 117$).

To answer the second research question – whether proficiency, operationalised as vocabulary size and listening skills, and aptitude, as measured by the LLAMA test, mediate the learning of vocabulary after sustained exposure to captioned video viewing – only data for the EGs ($n = 67$) were taken into consideration since the other participants were not exposed to the video viewing treatment (in line with Peters and Webb 2018). General Linear Models (GLMs) were run in SPSS using the Generalised Linear Mixed Model interface. In the models, listening skills, vocabulary size, LLAMA total score (as a comprehensive measure of participants' language aptitude) and level (high school vs. university) were entered as fixed effects. Subsequently, both GLMs were rerun without level as a fixed effect, since it was the only parameter which failed to reach statistical significance in all analyses (see Results section).

Results

Descriptive results computed for the pre- and post-tests (see Table 5) indicated that all participants, regardless of level and condition, knew significantly more of the target vocabulary forms and meanings at the end of the research period than at the beginning. These results were further analysed by paired samples *t*-tests and Wilcoxon signed-rank tests, based on the regularity conditions of the data. All of them confirmed that participants had made vocabulary learning gains over the academic term ($p < .001$ in all cases). In most cases, the EGs outperformed the CGs in both absolute and relative gains (see Table 6).

Table 5. Descriptive results of pre- and post-test in form and meaning per groups.

| | | Pre-test (%) | | | | Post-test (%) | | | |
| | | Form | | Meaning | | Form | | Meaning | |
Level	Group	Mean	SD	Mean	SD	Mean	SD	Mean	SD
High school	EG (*n* = 30)	17.60	9.75	3.19	5.21	39.56	19.42	18.39	15.51
	CG (*n* = 27)	16.19	10.76	2.43	4.93	32.98	18.35	10.53	8.96
University	EG (*n* = 37)	25.21	13.57	9.08	8.78	48.67	19.69	28.32	15.33
	CG (*n* = 23)	17.98	13.59	4.33	6.07	43.85	16.40	22.36	12.59

Table 6. Descriptive results of gains after treatment.

| | | Gains | Experimental Group | | | | Control Group | | | |
Level	Aspect	(%)	Mean	SD	Min.	Max.	Mean	SD	Min.	Max.
High school	Form	Absolute	21.96	13.57	0	45.00	16.78	11.29	0	37.50
		Relative	29.75	17.45	0	62.96	22.75	15.03	2.63	53.57
	Meaning	Absolute	15.20	13.00	0	45.00	8.09	6.75	0	22.50
		Relative	16.39	14.00	0	47.37	8.94	7.05	0	25.00
University	Form	Absolute	24.09	13.77	−12.50	52.50	25.87	10.68	10.00	48.57
		Relative	36.17	17.07	0	68.18	33.14	12.69	12.82	58.62
	Meaning	Absolute	19.62	10.18	0	45.00	18.02	9.51	5.00	40.00
		Relative	23.36	12.01	2.50	55.17	19.52	10.63	7.50	43.59

TW form

Research question 1: the impact of viewing captioned video on learning of vocabulary form
Results from the GLZ showed a significant main effect for condition ($F(1, 111) = 4.575$, $p = .035$) on the measures of TW form, confirming that participants in the EGs made significantly more vocabulary form gains than the CGs. There was also a significant effect for level on the number of TW forms

gained ($F(1, 111)= 8.413$, $p=.004$), with university students significantly outperforming high school learners. The interaction between the two fixed effects was not statistically significant ($F(1, 111)= .922$, $p= .339$). In order to know whether significant differences between conditions arose at both levels, simple contrasts between EGs and CGs at university and high school were considered. These revealed that differences between conditions were only statistically significant in high school ($\beta= 8.023$, $p= .033$), but not at university ($\beta= 4.035$, $p= .401$).

Research question 2: the mediating effect of vocabulary size, listening skills and language aptitude on learning of vocabulary form

In relation to the second research question, the first GLM showed a significant effect for listening skills ($F(1, 60)= 20.624$, $p=.000$), indicating that participants' listening proficiency influenced the learning of TW forms. The other factors did not have a significant effect on gains in knowledge of TW forms: vocabulary size ($F(1, 60)= 3.679$, $p=.060$); language aptitude ($F(1, 60)= .457$, $p= .502$) and level ($F(1, 60)= .161$, $p=.690$). However, in the second GLM run, without level as a fixed effect, both listening skills ($F(1, 61)= 20.766$, $p= .000$) and vocabulary size ($F(1, 61)=5.725$, $p= .020$) turned out to be statistically significant, supporting the idea that stronger listening proficiency and a larger vocabulary size favour the learning of TW forms. Language aptitude remained non-significant ($F(1, 61)= .398$, $p= .530$).

TW meaning recall

Research question 1: the impact of viewing captioned video on learning of vocabulary meaning

The GLZ revealed that the experimental condition played a significant role in the number of TW meanings gained during the intervention ($F(1, 109)= 8.154$, $p= .005$), with the EGs significantly outperforming the CGs. Level was also significant ($F(1, 109)= 16.003$, $p= .000$), with university students significantly outperforming high school students. However, the interaction between level and condition did not reach statistical significance ($F(1, 109)= 1.946$, $p= .166$). Simple contrasts between experimental conditions at each level were further examined and revealed that, in high school, the EG gained a significantly larger number of TW meanings than the CG ($\beta= 6.902$, $p= .005$), but this was not the case at university ($\beta= 3.840$, $p= .288$).

Research question 2: the mediating effect of vocabulary size, listening skills and language aptitude on learning of vocabulary meaning

Focusing on the mediating role of proficiency (as measured by vocabulary size and listening skills) and language aptitude on the meaning recall test, the first GLM showed that all fixed effects, except for level ($F(1, 60)= 3.220$, $p=.078$), played a significant role: listening skills ($F(1, 60)= 12.403$, $p= .001$), vocabulary size ($F(1, 60)= 6.101$, $p= .016$), and language aptitude ($F(1, 60)= 5.591$, $p= .021$). In the second model that was run, without the non-significant factor (i.e. level), the same three variables remained statistically significant: vocabulary size ($F(1, 61)= 13.424$, $p= .001$), listening skills ($F(1, 61)= 10.724$, $p= .002$) and language aptitude ($F(1, 61)= 4.411$, $p= .040$), showing that the combination of stronger listening skills and higher language aptitude, together with a larger vocabulary size, favoured the learning of TW meanings.

Discussion

The aim of this study was twofold: first, to examine the impact on vocabulary learning of captioned video viewing in support of formal classroom instruction, compared with formal classroom instruction only; second, to investigate the extent to which proficiency and language aptitude mediated the impact of captioned video on EG participants' vocabulary learning. Results revealed that all learners made significant gains in their knowledge of TWs over the course of the intervention, as shown by the paired-samples *t*-tests and Wilcoxon signed-rank tests. Results from the GLZs showed that both level and experimental condition significantly influenced the gains in knowledge of TW

forms and meanings. However, viewing captioned video only played a significant role in high school, not at university. Regarding the second research question, the GLMs revealed that both listening skills and vocabulary size significantly influenced the extent of learners' gains in knowledge of TW forms and meanings, while language aptitude only proved to be statistically significant in the learning of TW meanings.

TW form

The results from the spelling test show that significant word form learning occurred in both conditions. This is understandable since deliberate attention to language items can lead to learning if learners become aware of the target language input (Schmidt 2010). In the present study, the TW words were presented at the beginning of each session through a range of exercises and then encountered again on the post-tasks. In this way, participants' attention was directed towards the target vocabulary and they could employ previously developed language learning strategies to do their utmost to learn them.

Participants' level of study influenced the learning of TW forms as the university group learned comparatively more than the high school group; this is in line with previous research that found a positive effect for proficiency on vocabulary learning from TV viewing (e.g. Frumuselu 2015; Peters and Webb 2018). However, the extra exposure to the word forms provided by viewing captioned video was significantly beneficial not for the university group, but for the high school students. This advantage can be explained with reference to the higher proficiency level of the university students and their greater language learning experience. Learners at lower levels (i.e. the high-schoolers) may be assumed to have less developed FL skills and language learning strategies compared to the more advanced learners (i.e. the university group). It could be argued that viewing captioned video would be particularly beneficial for less proficient students because it would allow them to hear and 'see' the target vocabulary in context and would provide additional input for learning. L2 subtitles in particular would help them match the aural and written forms of words (Borrás and Lafayette 1994; Peters, Heynen, and Puimège 2016). In this sense, the exposure to the TWs delivered at the beginning and the end of classes, which was the only input received by the CGs, could be seen as less effective for learning for the less proficient learners. In the case of the more proficient university learners, extra exposure to the TWs through captioned video viewing may have been unnecessary; the formal instruction and teacher feedback received by both EG and CG could have been sufficient to trigger their learning of the TWs. Certainly, as in other studies (Laufer 2005; Schmitt 2008), there is evidence to suggest that intentional learning can lead to greater and faster vocabulary gains in the L2, at least at more advanced proficiency levels.

It has to be acknowledged that the number of word forms learned could have been higher for all conditions: the EGs gained 32.96% of TW forms and the CGs 27.95%; in other words, 13 word forms in the EG and 11 in the CG out of a total of 40. Seventy five percent of the TWs appeared between two and four times in the eight video episodes shown; it may be that more encounters are needed to trigger greater word form learning, as has been suggested by previous research (Rott 1999; Waring and Takaki 2003; Webb 2007).

Regarding possible mediating factors influencing vocabulary learning supported by captioned video viewing, our study confirms a role for proficiency in TW form gains, as established by existing research (Montero Perez et al. 2014; Peters and Webb 2018). The GLM results suggest that the higher a student's vocabulary size and OPT listening score, the greater the gains in TW form knowledge when exposed to captioned video viewing. In connection to this, Webb and Rodgers (2009) found that knowledge of 3,000 word families was enough to enable successful television viewing; specifically, they claimed that in older programmes (the category in which they included two episodes of *I Love Lucy*), knowledge of the first three thousand words allowed for 96.26% of lexical coverage, which has been said to be enough for reasonable comprehension of a text (Laufer 1989). In the

present study, 95% lexical coverage was achieved at the 2 K level, indicating that knowledge of the most frequent two thousand words of the English language was enough to understand the TV episodes selected. EG learners had a mean VS of 3,455 (high school group) and 4,528 words (university group), which should have enabled them to understand the audiovisual materials reasonably well. Our argument is that students with a higher VS are likely to be better able to concentrate on new (unknown) vocabulary, such as the TWs in this study, and allocate greater attentional resources and learning strategies to decipher their aural and written forms. This explanation can be extended to explain the significant role that listening skills appeared to play in the learning of TW forms. Participants' developed L2 listening skills would have allowed them to parse speech and better isolate the word forms they were exposed to. Of course, if this did not suffice in the case of more challenging vocabulary, participants could still resort to the captions and the written forms of the target vocabulary, engaging their reading rather than listening skills.

Aptitude, however, was not found to be relevant in TW form gains, based on the GLM results. This could be due to the way the participants approached the task of learning the TWs. This learning task was presented in a highly explicit, teacher-led fashion both for the EG and the CG, a format that all participants would have been familiar with. Participants may have simply engaged relatively superficial learning strategies such as memorisation, note-taking or selective attention, thus focusing on the TWs only. Although the overall learning gains were not impressive as compared to the total number of words that could have been learned, it might be the case that the learning experience during the intervention was insufficiently challenging, as students were presented only five words in each session. We therefore suggest that higher proficiency, wider learning experience, and other learning mechanisms, abilities and strategies could be diminishing the impact of aptitude, as they have been found to do in the case of other cognitive abilities (Robinson 2001), such as working memory (Gilabert et al. 2016). Thus, it would appear that students at this level of proficiency and with this video input did not need to draw so much on language learning aptitude to process forms of novel vocabulary, though they might have had to if the target vocabulary and the content of the videos had been more demanding (relative to their proficiency level).

TW meaning recall

The results of the GLZ on the meaning recall test also showed a significant effect for level. It should be noted that the content of *I Love Lucy* was probably more suitable for the university students, regardless of their higher proficiency. Although this TV show is a situation comedy with recurrent clichés and jokes, the high school students may have needed more scaffolding and content support to understand the storyline and to adjust their worldview to the one shown in the series. This could have added to the overall multimodal cognitive load (Mayer, Lee, and Peebles 2014) and, therefore, may have played against their vocabulary learning. The less proficient participants may have found the novel vocabulary less accessible because they were more focused on processing other aspects of the video, such as external references or grammatical forms, besides the images and the captions. In contrast, the university students, who we assume were able to understand the content of the video more easily, may have been better able to direct more of their mental capacity to attending to the novel vocabulary.

The significant differences between the experimental and the control groups concerning TW meaning recall confirm other research which has found that the co-occurrence of TWs with associated on-screen imagery can support vocabulary learning, especially of low-frequency vocabulary (Rodgers 2018). In our study, learners in the EGs benefitted from on-screen exposure to the TWs, 75% of which were mid- or low-frequency according to Nation (2013), but expressed mostly concrete concepts, following Brysbaert, Warriner, and Kuperman (2014) (see Appendix). That learners could often see a visual representation of the concept while being exposed to the aural and written form of the word could be argued to be particularly supportive to learning meaning. This argument is in line with Mayer's multimedia principle, according to which 'people learn better from words and

picture than from words alone' (Mayer 2009: 4). When simultaneously exposed to words and images, learners are able to build verbal and visual mental representations and establish connections between the two, which leads to greater depth of processing and meaningful learning. When presented with words alone, learners can only build a verbal mental model, and it is less likely that they will build a visual one, such that connections between the two systems are less likely.

As noted in the case of TW form, the meaning recall test results suggest that the less proficient group benefitted more from hearing and seeing the TWs in context, and from the visual support provided by the video. This is in line with Mayer's (2009: 223) suggestion that the multimedia principle 'may apply more strongly to low-knowledge learners than to high-knowledge learners, presumably because low-knowledge learners need guidance in building referential connections between pictorial and verbal representations'.

Again, it has to be acknowledged that the number of word meanings learned could have been greater for the experimental conditions: the EGs showed a gain of 19.88% or an average of around 8 word meanings out of 40 per student, compared with the CGs' gain of 14.23%, around six word meanings per student. Nevertheless, these gains are in line with those reported in other studies in the field. Peters and Webb (2018), for instance, found gains of 8.31% in the EG and 3.35% in the CG on the meaning recall test, while Rodgers (2013) found that the EG gained 26.20% and the CG 23.14% of the TW meanings on average, based on the two tests (tough and sensitive) used. However, both these studies differed significantly from ours in that they did not involve any formal explicit instruction of the TWs. In addition, Peters and Webb's EG did not view the video with captions, while Rodgers used a meaning recognition test, rather than a meaning recall test, as in our case.

As with the TW form results, proficiency was found to mediate TW meaning learning supported by captioned video viewing. Our results concerning the facilitating effect of vocabulary size confirm those from previous research on vocabulary learning and video viewing (Montero Perez et al. 2014; Montero Perez, Peters, and Desmet 2015). As argued by Na and Nation (1985), larger vocabulary size helps learners derive word meanings from context. There are several reasons for this: firstly, the encounters of unknown words will be fewer; secondly, new contexts will be easier to interpret with consolidated knowledge of the vocabulary co-occurring with the unknown words; and thirdly, the learning burden of new words will be lessened thanks to learners' greater mastery of the rules of the language, and hence word meanings will be more easily and rapidly learned (Webb and Nation 2017).

The GLM also showed that the stronger learners' listening skills, the greater the gains in TW meanings. As the 'quality of a learner's listening comprehension is strongly dependent on his ability to cope with the heavy on-line processing demands of understanding spoken input' (Staehr 2008: 148), it stands to reason that stronger listening skills should lead to better comprehension, which involves understanding word meaning. With stronger listening skills, learners are better able to pick up new words and work out their meaning, or consolidate their understanding of partially known vocabulary. Better listening skills can also help L2 listeners to distinguish vocabulary that they might otherwise fail to identify (van Zeeland 2013). As was the case for learning word forms, this will foster their segmentation abilities, which will also contribute to better listening comprehension and thus facilitate word meaning learning.

Unlike the results relating to TW form gains, TW meaning gains after captioned video viewing were influenced by aptitude scores. As explained above, three out of the four LLAMA sub-tests (LLAMA B, E and F) tap into explicit language aptitude. The learners in this study were generally used to focusing on word meaning explicitly as part of their formal language learning, and during the pre-teaching associated with this study, explicit focus was drawn more to TW meaning than to form. While learning the TW forms was no doubt a more straightforward procedure, possibly engaging superficial learning strategies rather than any language aptitude, learning TW meaning may have involved a greater cognitive involvement where language learning aptitude may have come into play.

Conclusion, limitations and further research

Our findings relating to vocabulary learning supported by captioned video viewing are consistent with previous research, showing that captioned video can provide an effective support for instructed vocabulary learning, particularly for less proficient learners. It has been found that using TV series in the classroom is not detrimental to learning, as the CGs did not significantly out-perform the EGs. However, the CGs in this study also proved to benefit from the guided exposure to the target vocabulary. Therefore, it seems sensible that, if some video viewing activity is implemented in the classroom, this should be complemented with some focused activities so that all students make the most of the video viewing experience. It might also be recommendable to show the videos with textual support, as they seem to especially facilitate the learning of TW forms. Without these two conditions (focused video viewing and captions), it is likely that the gains would have been lower and the intervention less beneficial. Finally, this in-class rich learning experience from TV viewing might also have the potential to guide more informal, out-of-class learning (Webb 2015), in which vocabulary may be learned incidentally given the right number of encounters.

This study also suggests that higher proficiency, and aptitude to a certain extent, matter in vocabulary learning in relation to the support offered by captioned video viewing. If video content is demanding (relative to the students' proficiency), novel vocabulary learning is less likely, but higher aptitude could play a role here. If video content is easier (relative to one's proficiency), novel vocabulary learning is more likely, and aptitude may not be a factor in whether a learner can use the video input to enhance their vocabulary learning. Thus, four factors – the video content difficulty, the amount of novel vocabulary in the video, the proficiency level of the learners, and their language learning aptitude – may interact, counterbalancing or counter-acting each other. When the four factors are aligned at optimum level, there is strong potential for successful new vocabulary learning.

This study is not without limitations. For instance, some of the characteristics of the TWs could not be totally controlled for (e.g. part of speech, frequency of occurrence or word length) as it is almost impossible to control for these factors when using authentic audiovisual materials in the classroom. However, these authentic materials guarantee the ecological validity of this research. Another limitation lies in the inclusion of a final focus on the TWs at the end of each session as the students' attention was drawn to them. This might have had an effect on learners' processing of the TWs and so could have increased learning gains. However, as both groups went through the same procedure (except for the EG's extra exposure to the TWs in the captioned video), the effects of the processing of the TWs should logically have been the same for both experimental conditions. Further, as the study took place in 'real' classes, rather than in a laboratory setting, it was not possible to check in any way whether the participants in the EGs were actually reading the captions or not when viewing the video, and thus, whether they in fact received additional exposure to the written forms of the TWs. Eye-tracking could be used to check this in an experimental setting. Eye-tracking studies have already confirmed that learners do tend to read captions, which benefits later word recognition (Montero Perez, Peters, and Desmet 2015), though their eye behaviour is conditioned by their age and proficiency (Muñoz 2017), the participants' L1 (Winke, Gass, and Sydorenko 2013) and the genre of the videos (Gilabert et al. forthcoming).

Regarding further research, a delayed post-test would shed more light on the actual learning and retention of the TWs in the long term. In addition, although we have referred to the argument that co-occurrence of a TW (aurally and its written form in captions) and its visual representation on-screen supports the learning of TW meaning, it should be noted that this idea was not analysed in the video episodes used in this study. It would therefore be interesting to analyse the extent to which the visual and verbal representations of the forty TWs co-occurred, and to investigate whether this had any association with participants' learning.

Notes

1. Named after the Gospel according to Matthew in the New Testament: 'For to those who have, more will be given, and they have an abundance; but from those who have nothing, even what they have will be taken away' (Matthew 13:12, New Revised Standard Version Bible).
2. By *teacher-led instruction* we understand pre-teaching the TW forms and meanings through a focus-on-forms vocabulary task at the beginning of the class (for both experimental and control groups) and the administration of a vocabulary post-task after viewing the episode (experimental group) or at the end of class (control group).
3. Throughout the paper, *multimodal input* is understood as the simultaneous presentation of L2 text (subtitles), L2 sound (soundtrack) and video image.
4. In this paper, *L2 subtitles* and *captions* will be used interchangeably.
5. Interestingly, some aptitude component tests which, per se, might not be expected to be related to vocabulary - e.g. the Language Analytic Ability test of the Pimsleur Language Aptitude test (PLAB) (Pimsleur, Reed, and Stansfield 2004) - have been found to associate with differences in L2 vocabulary size, though not in L1 vocabulary size (Dąbrowska 2019).
6. If all the meanings had been accepted on the pre- and post-tests, it would not have been possible to know whether participants learned them thanks to the TV series or for any other reason.
7. For this correspondence, only the scores of the X_Lex test were taken into account as Meara and Milton (2003), Milton (2010) and Milton and Alexiou (2009) did not administer the Y_Lex test. As seen in Table 3, the total score of both vocabulary size tests was 4,415 words for the university group and 3,475 for high-schoolers.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Maria del Mar Suárez http://orcid.org/0000-0002-1741-7596
Ferran Gesa http://orcid.org/0000-0003-1775-3212

References

Alexiou, T. 2015. Vocabulary uptake from *Peppa Pig*: a case study of preschool EFL learners in Greece. In *Current Issues in Second / Foreign Language Teaching and Teacher Development: Research and Practice*, ed. C. Gitsaki and T. Alexiou, 285–301. Newcastle upon Tyne: Cambridge Scholars Publishing.
Allan, D. 2004. *Oxford Placement Test 1*. Oxford: Oxford University Press.
Andringa, S. and P. Rebuschat. 2015. New directions in the study of implicit and explicit learning. *Studies in Second Language Acquisition* 37, no. 2: 185–196. DOI: 10.1017/S027226311500008X.
Aurstad, L.M.G. 2013. The role of subtitles in second language acquisition: an experimental study in the context of the Norwegian school system. MA diss. Norwegian University of Science and Technology.
Bolibaugh, C. and P. Foster. 2013. Memory-based aptitude for nativelike selection: the role of phonological short-term memory. In *Sensitive Periods, Language Aptitude, and Ultimate L2 Attainment*, ed. G. Granena and M. Long, 205–230. Philadelphia: John Benjamins.
Borrás, I. and R.C. Lafayette. 1994. Effects of multimedia courseware subtitling on the speaking performance of college students of French. *The Modern Language Journal* 78, no. 1: 61–75. DOI: 10.2307/329253.

Bravo, M.C.C. 2008. Putting the reader in the picture: screen translation and foreign-language learning. PhD diss. Universitat Rovira i Virgili.

Brysbaert, M., A.B. Warriner and V. Kuperman. 2014. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods* 46, no. 3: 904–911. DOI: 10.3758/s13428-013-0403-5.

Carroll, J.B. 1981. Twenty-five years of research on foreign language aptitude. In *Individual Differences and Universals in Language Learning Aptitude*, ed. K.C. Diller, 83–118. Rowley, MA: Newbury House.

Carroll, J.B. 1993. *Human Cognitive Abilities: A survey of Factor-analytic Studies*. Cambridge, UK: Cambridge University Press.

Carroll, J.B. and S.M. Sapon. 1959. *Modern Language Aptitude Test*. San Antonio, TX: Psychological Corporation.

Carroll, J.B. and S.M. Sapon. 2002. *Modern Language Aptitude Test: Manual*, second edition. Bethesda, MD: Second Language Testing, Inc.

Cobb, T. ongoing. *Compleat Web VocabProfile*. (version 2.0).https://www.lextutor.ca/vp/comp/ (accessed June 12, 2018).

Cokely, M. and C. Muñoz. forthcoming. Captioned video, vocabulary and visual prompts: an exploratory study. In *Using Film and Media in the Language Classroom: Reflections on Research-led Teaching*, ed. C. Herrero and I. Vanderschelden, 61–75. Bristol: Multilingual Matters.

Coxhead, A., P. Nation and D. Sim. 2015. Measuring the vocabulary size of native speakers of English in New Zealand secondary school. *New Zealand Journal of Educational Studies* 50, no. 1: 121–135. DOI: 10.1007/s40841-015-0002-3.

Dąbrowska, E. 2019. Experience, aptitude, and individual differences in linguistic attainment: a comparison of native and nonnative speakers. *Language Learning* 69, no. S1: 72–100. DOI: 10.1111/lang.12321.

Dahlen, K. and C. Caldwell-Harris. 2013. Rehearsal and aptitude in foreign vocabulary learning. *The Modern Language Journal* 97, no. 4: 902–916. DOI: 10.1111/j.1540-4781.2013.12045.x.

Davies, M. 2009. The 385+ million word *Corpus of Contemporary American English* (1990–2008+). Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics* 14, no. 2: 159–190. DOI: 10.1075/ijcl.14.2.02dav.

Doughty, C. 2019. Cognitive language aptitude. *Language Learning* 69, no. S1: 101–126. DOI: 10.1111/lang.12322.

Elley, W.B. 1989. Vocabulary acquisition from listening to stories. *Reading Research Quarterly* 24, no. 2: 174–187. http://www.jstor.org/stable/747863.

Etemadi, A. 2012. Effects of bimodal subtitling of English movies on content comprehension and vocabulary recognition. *International Journal of English Linguistics* 2, no. 1: 239–248. DOI: 10.5539/ijel.v2n1p239.

European Commission, Directorate-General for Education, Youth, Sport and Culture. 2011. *Study on the Use of Subtitling. The Potential of Subtitling to Encourage Foreign Language Learning and Improve the Mastery of Foreign Languages* (Report No. EACEA/2009/01). https://publications.europa.eu/en/publication-detail/-/publication/e4d5cbf4-a839-4a8a-81d0-7b19a22cc5ce.

Frumuselu, A.D. 2015. Subtitled television series inside the EFL classroom: long-term effects upon colloquial language learning and oral production. PhD diss., Universitat Rovira i Virgili.

Galimberti, V. 2016. Watching TV series with L1 or L2 subtitles: a study on comprehension and vocabulary acquisition with Italian beginner learners of English. MA diss., Universitat de Barcelona.

Gesa, F. forthcoming. L1/L2 subtitled videos and EFL learning: a study on vocabulary acquisition and content comprehension at different proficiency levels. PhD diss., Universitat de Barcelona.

Ghia, E. 2012. *Subtitling Matters: New Perspectives on Subtitling and Foreign Language Learning*. Bern: Peter Lang.

Gilabert, R., J.C. Mora, C. Muñoz, and F. Gesa. 2016. Individual differences in the learning of L2 vocabulary through captioned video. Paper presented at the 35th Second Language Research Forum, New York, NY, 22–25 September, 2016.

Gilabert, R., M.M. Suárez, N. Moskvina, M. Levkina, O. Vasylets, and S. Feijoo. forthcoming. The impact of genre on L2 vocabulary learning and eye-behaviour through captioned video.

Granena, G. 2013. Cognitive aptitudes for L2 learning and the LLAMA language aptitude test. In *Sensitive Periods, Language Aptitude, and Ultimate Attainment*, ed. G. Granena and M. Long, 179–204. Philadelphia: John Benjamins.

Granena, G. and M.C. Long. 2013. Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research* 29, no. 3: 311–343. DOI: 0.1177/0267658312461497.

Horst, M., T. Cobb and P. Meara. 1998. Beyond a clockwork orange: acquiring second language vocabulary through reading. *Reading in a Foreign Language* 11, no. 2: 207–223. http://nflrc.hawaii.edu/rfl/PastIssues/rfl112horst.pdf.

Hsu, W. 2013. The effects of audiovisual support on EFL learners' productive vocabulary. *ReCALL* 26, no. 1: 62–79. DOI: 10.1017/S0958344013000220.

Hwu, F. and S. Sun. 2012. The aptitude-treatment interaction effects on the learning of grammar rules. *System* 40, no. 4: 505–521. DOI: 10.1016/j.system.2012.10.009.

Kaufman, S.B., C.G. DeYoung, J.R. Gray, L. Jiménez, J. Brown and N. Mackintosh. 2010. Implicit learning as an ability. *Cognition* 116, no. 3: 321–340. DOI: 10.1016/j.cognition.2010.05.011.

Kormos, J. and A. Sáfár. 2008. Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition* 11, no. 2: 261–271. DOI: 10.1017/S1366728908003416.

Kvitnes, I.E.N. 2013. Subtitles in the second language classroom: an experimental study with Norwegian learners of English. MA diss., Norwegian University of Science and Technology.

Laufer, B. 1989. What percentage of text lexis is essential for comprehension? In *Special Language: From Humans Thinking to Thinking Machines*, ed. C. Laurén and M. Nordman, 316–323. Clevedon, UK: Multilingual Matters.

Laufer, B. 2005. Focus on form in second language vocabulary learning. *EUROSLA Yearbook* 5: 223–250. DOI: 10.1075/ eurosla.5.11lau.

Laufer, B. 2006. Comparing focus on form and focus on formS in second-language vocabulary learning. *The Canadian Modern Language Review/La Revue Canadienne des Langues Vivantes* 63, no. 1: 149–166. DOI: 10.1353/cml.2006.0047.

Laufer, B. and G.C. Ravenhorst-Kalovski. 2010. Lexical threshold revisited: lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language* 22, no. 1: 15–30. http://nflrc.hawaii.edu/rfl/April2010/ articles/laufer.pdf.

Li, S. 2015. The associations between language aptitude and second language grammar acquisition: a meta-analytic review of five decades of research. *Applied Linguistics* 36, no. 3: 385–408. DOI: 0.1093/applin/amu054.

Li, S. 2016. The construct validity of language aptitude: a meta-analysis. *Studies in Second Language Acquisition* 38, no. 4: 801–842. DOI: 10.1017/S027226311500042X.

Lindgren, E. and C. Muñoz. 2013. The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism* 10, no. 1: 105–129. DOI: 10.1080/14790718. 2012.679275.

Malone, J. 2018. Incidental vocabulary learning in SLA: effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition* 40, no. 3: 651–675. DOI: 10.1017/S0272263117000341.

Maneshi, N. 2017. Incidental vocabulary learning through listening to songs. MA diss., The University of Western Ontario.

Mayer, R.E. 2009. *Multimedia Learning*, second edition. Cambridge, UK: Cambridge University Press.

Mayer, R.E., H. Lee and A. Peebles. 2014. Multimedia learning in a second language: a cognitive load perspective. *Applied Cognitive Psychology* 28, no. 5: 653–660. DOI: 10.1002/acp.3050.

Meara, P. 2005a. *LLAMA Language Aptitude Tests*. Swansea: Lognostics.

Meara, P. 2005b. *X_Lex: The Swansea Vocabulary Levels Test* (version 2.05). Swansea: Lognostics.

Meara, P. and J. Milton. 2003. *X_Lex. The Swansea Levels Test*. Newbury: Express.

Meara, P. and I. Miralpeix. 2006. *Y_Lex: The Swansea Advanced Vocabulary Levels Test* (version 2.05). Swansea: Lognostics.

Milton, J. 2010. The development of vocabulary breadth across the CEFR levels. A common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, and textbooks across Europe. In *EuroSLA Monograph Series 1 - Communicative Proficiency and Linguistic Development: Intersections Between SLA and Language Testing Research*, ed. I. Bartning, M. Martin and I. Vedder, 211–232. http://eurosla.org/monographs/EM01/211-232Milton.pdf.

Milton, J. and T. Alexiou. 2009. Vocabulary size and the common European framework of reference for languages. In *Vocabulary Studies in First and Second Language Acquisition: The Interface between Theory and Application*, ed. B. Richards, M.H. Daller, D.D. Malvern, P. Meara, J. Milton and J. Treffers-Daller, 194–211. London: Palgrave Macmillan.

Miralpeix, I. 2012. X_Lex and Y_Lex: A validation study. Paper presented at the 22nd Lexical Studies Conference, 16–19 March, 2012, Newtown, UK.

Miralpeix, I. and C. Muñoz. 2018. Receptive vocabulary size and its relationship to EFL language skills. *IRAL - International Review of Applied Linguistics in Language Teaching* 56, no. 1: 1–24. DOI: 10.1515/iral-2017-0016.

Mitterer, H. and J.M. McQueen. 2009. Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS ONE* 4, no. 11: 1–5. DOI: 10.1371/journal.pone.0007785.

Miyake, A. and N.P. Friedman. 1998. Individual differences in second language proficiency: working memory as language aptitude. In *Foreign Language Learning: Psycholinguistic Studies on Training and Retention*, ed. A.F. Healy and L.E. Bourne, 339–364. Mahwah, NJ: Erlbaum.

Montero Perez, M., E. Peters, G. Clarebout and P. Desmet. 2014. Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology* 18, no. 1: 118–141. DOI: 10125/44357.

Montero Perez, M., E. Peters and P. Desmet. 2015. Enhancing vocabulary learning through captioned video: an eye-tracking study. *The Modern Language Journal* 99, no. 2: 308–328. DOI: 10.1111/modl.12215.

Montero Perez, M., E. Peters and P. Desmet. 2018. Vocabulary learning through viewing video: the effect of two enhancement techniques. *Computer Assisted Language Learning* 31, no. 1–2: 1–26. DOI: 10.1080/09588221.2017.1375960.

Montero Perez, M., W. Van den Noortgate and P. Desmet. 2013. Captioned video for L2 listening and vocabulary learning: a meta-analysis. *System* 41, no. 3: 720–739. DOI: 10.1016/j.system.2013.07.013.

Muñoz, C. 2008. Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics* 29, no. 4: 578–596. DOI: 10.1093/applin/amm056.

Muñoz, C. 2017. The role of age and proficiency in subtitle reading. An eye-tracking study. *System* 67, no. 3: 77–86. DOI: 10.1016/j.system.2017.04.015.

Na, L. and P. Nation. 1985. Factors affecting guessing vocabulary in context. *RELC Journal* 16, no. 1: 33–42. DOI: 10.1177/ 003368828501600103.

Nation, P. 2006. How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review/La Revue Canadienne des Langues Vivantes* 63, no. 1: 59–82. DOI: 10.3138/cmlr.63.1.59.

Nation, P. 2012. The BNC/COCA Word family lists. *Victoria University of Wellington*, September 17. https://www.victoria.ac. nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf (accessed 13 June, 2018).

Nation, P. 2013. *Learning Vocabulary in Another Language*, second edition. Cambridge, UK: Cambridge University Press.

Nation, P. and R. Waring. 1997. Vocabulary size, text coverage and word lists. In *Vocabulary: Description, Acquisition and Pedagogy*, ed. N. Schmitt and M. McCarthy, 6–19. Cambridge: Cambridge University Press.

Nation, P. and S. Webb. 2011. *Researching and Analyzing Vocabulary*. Boston: Heinle.

Noreillie, A.-S., B. Kestemont, K. Heylen, P. Desmet and E. Peters. 2018. Vocabulary knowledge and listening comprehension at an intermediate level in English and French as foreign languages. An approximate replication study of Staehr (2009). *ITL - International Journal of Applied Linguistics* 169, no. 1: 212–231. DOI: 10.1075/itl.00013.nor.

Oppenheimer, J. and D. Arnaz. 1951. *I Love Lucy*, Season 5 DVD. Directed by J.V. Kern. Los Angeles: Desilu Productions.

Penno, J.F., I.A.G. Wilkinson and D.W. Moore. 2002. Vocabulary acquisition: from teacher explanation and repeated listening to stories: do they overcome the Matthew effect? *Journal of Educational Psychology* 94, no. 1: 23–33. DOI: 10.1037/0022-0663.94.1.23.

Peters, E., E. Heynen and E. Puimège. 2016. Learning vocabulary through audiovisual input: the differential effect of L1 subtitles and captions. *System* 63: 134–148. DOI: 10.1016/j.system.2016.10.002.

Peters, E. and S. Webb. 2018. Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition* 40, no. 3: 551–577. DOI: 10.1017/S0272263117000407.

Pimsleur, P., D.J. Reed and C. Stansfield. 2004. *Pimsleur Language Aptitude Battery: Manual*. Bethesda, MD: Second Language Testing, Inc.

Plonsky, L. 2015. Sample demographics and generalizability in SLA: A systematic review. Paper presented at the Workshop on Reliability and validity in SLA research, 11–12 June 2015, Barcelona, Spain.

Robinson, P. 2001. Individual differences, cognitive abilities, aptitude complexes, and learning conditions in SLA. *Second Language Research* 17, no. 4: 368–392.

Rodgers, M.P.H. 2013. *English language learning through viewing television: an investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions*. PhD diss., Victoria University of Wellington.

Rodgers, M.P.H. 2018. The images in television programs and the potential for learning unknown words. The relationship between on-screen imagery and vocabulary. *ITL - International Journal of Applied Linguistics* 169, no. 1: 191–211. DOI: 10.1075/itl.00012.rod.

Rogers, V., P. Meara, T. Barnett-Legh, C. Curry and E. Davie. 2017. Examining the LLAMA aptitude tests. *Journal of the European Second Language Association* 1, no. 1: 49–60. DOI: 10.22599/jesla.24.

Rott, S. 1999. The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition* 21, no. 4: 589–619. DOI: 10.1017/S0272263199004039.

Saito, K. 2017. Effects of sound, vocabulary and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning* 67, no. 3: 665–693. DOI: 10.1111/lang.12244.

Schmidt, R. 2010. *Attention, awareness, and individual differences in language learning*. Proceedings of the Fourth Centre for Language Studies International Conference, Singapore, 2010, 721–737. Singapore: National University of Singapore, Centre for Language Studies.

Schmitt, N. 2008. Instructed second language vocabulary learning. *Language Teaching Research* 12, no. 3: 329–363. DOI: 10.1177/1362168808089921.

Serafini, E.J. and C. Sanz. 2016. Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition* 38, no. 4: 607–646. DOI: 10.1017/S0272263115000327.

Skehan, P. 2012. Language aptitude. In *Handbook of Second Language Acquisition*, ed. S. Gass and A. Mackey, 381–395. New York, NY: Routledge.

Staehr, L.S. 2008. Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal* 36, no. 2: 139–152. DOI: 10.1080/09571730802389975.

Staehr, L.S. 2009. Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition* 31, no. 4: 577–607. DOI: 10.1017/S0272263109990039.

Stanovich, K.E. 2009. Matthew effects in reading: some consequences of individual differences in the acquisition of literacy. *Journal of Education* 189, no. 1–2: 23–55. DOI: 10.1177/0022057409189001-204.

Sydorenko, T. 2010. Modality of input and vocabulary acquisition. *Language Learning & Technology* 14, no. 2: 50–73. DOI: 10125/44214.

Vanderplank, R. 2010. *Déjà Vu*? A decade of research on language laboratories, television and video in language learning. *Language Teaching* 43, no. 1: 1–37. DOI: 10.1017/S0261444809990267.

Vanderplank, R. 2016. *Captioned Media in Foreign Language Learning and Teaching: Subtitles for the Deaf and Hard-of-Hearing as Tools for Language Learning*. London: Palgrave Macmillan.

van Zeeland, H. 2013. Vocabulary and listening. In *The Encyclopedia of Applied Linguistics*, ed. C.A. Chapelle. Chichester, UK: Wiley-Blackwell.

van Zeeland, H. and N. Schmitt. 2013. Incidental vocabulary acquisition through L2 listening: a dimensions approach. *System* 41, no. 3: 609–624. DOI: 10.1016/j.system.2013.07.012.

Vidal, K. 2003. Academic listening: A source of vocabulary acquisition? *Applied Linguistics* 24, no. 1: 56–89. DOI: 10.1093/applin/24.1.56.

Vidal, K. 2011. A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning* 61, no. 1: 219–258. DOI: 10.1111/j.1467-9922.2010.00593.x.

Waring, R. and M. Takaki. 2003. At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language* 15, no. 2: 130–160. http://nflrc.hawaii.edu/rfl/October2003/waring/waring.pdf.

Webb, S. 2007. The effects of repetition on vocabulary knowledge. *Applied Linguistics* 28, no. 1: 46–65. DOI: 10.1093/applin/aml048.

Webb, S. 2015. Extensive viewing: language learning through watching television. In *Language Learning Beyond the Classroom*, ed. D. Nunan and J.C. Richards, 159–168. New York: Routledge.

Webb, S. and A.C.-S. Chang. 2015. How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition* 37, no. 4: 651–675. DOI: 10.1017/S0272263114000606.

Webb, S. and P. Nation. 2017. *How Vocabulary is Learned*. Oxford: Oxford University Press.

Webb, S. and M.P.H. Rodgers. 2009. Vocabulary demands of television programs. *Language Learning* 59, no. 2: 335–366. DOI: 10.1111/j.1467-9922.2009.00509.x.

Williams, J.N. 2009. Implicit learning in second language acquisition. In *The New Handbook of Second Language Acquisition*, ed. W.C. Ritchie and T.K. Bhatia, 319–353. Bingley, UK: Emerald Group.

Winke, P. 2013. An investigation into second language aptitude for advanced Chinese learning. *The Modern Language Journal* 97, no. 1: 109–130. DOI: 10.1111/j.1540-4781.2013.01428.x.

Winke, P., S. Gass and T. Sydorenko. 2010. The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology* 14, no. 1: 65–86. DOI: 10125/44203.

Winke, P., S. Gass and T. Sydorenko. 2013. Factors influencing the use of captions by foreign language learners: an eye-tracking study. *The Modern Language Journal* 97, no. 1: 254–275. DOI: 10.1111/j.1540-4781.2013.01432.x.

Yuksel, D. and B. Tanriverdi. 2009. Effects of watching captioned movie clip on vocabulary development of EFL learners. *The Turkish Online Journal of Educational Technology* 8, no. 2: 48–54. http://tojet.net/articles/v8i2/824.pdf.

Zhang, X. 2013. The I don't know option in the vocabulary size test. *TESOL Quarterly* 47, no. 4: 790–811. DOI: 10.1002/tesq.98.

Appendix. Characteristics of the TWs tested

TWs	Frequency		COCA* (Freq. band)	Part of speech	Concreteness** (Mean)	Cognateness (Y / N)
	Episode (N times)	Term (N times)				
Beady eyes	2	2	4k	N	3.46	N
Bucket	5	5	2k	N	4.96	N
Cheapskate	2	2	17k	Adj.	3.26	N
Choppy	3	3	11k	Adj.	2.92	N
Conductor	4	4	3k	N	4.26	N
Crowbar	3	5	12k	N	4.87	N
Crummy	4	4	14k	Adj.	2.57	N
Cue tip	2	2	5k	N	3.31	N
Cuffs	5	5	6k	N	4.61	N
Curler	4	4	2k	N	4.13	N
Downtown	4	4	Off-list	N	4.39	N
Dummy	3	3	6k	N	3.96	N
To fool	3	5	2k	V	3.19	N
Forecourt	6	6	13k	N	–	N
To forge	4	4	4k	V	4.04	N
Furnace	2	3	7k	N	4.69	N
Galoshes	4	4	16k	N	–	N
Gear	2	2	2k	N	4.28	N
Grapefruit	2	3	9k	N	4.96	N
To hobnob	2	2	16k	V	–	N
Hunk	2	3	9k	N	3.79	N
Mitten	3	3	10k	N	4.89	N
Newsstand	2	2	Off-list	N	4.71	N
To peek through	5	5	7k	V	3.62	N
Penthouse	5	5	10k	N	4.36	N
Preview	4	7	6k	N	3.20	Y
Razor	3	3	6k	N	4.90	N
To roll	6	7	1k	V	4.16	Y
Rubdown	4	4	Off-list	N	3.62	N
Rugged	3	3	6k	Adj.	2.75	N
Seasick	2	2	Off-list	Adj.	3.18	N
Sneaky	2	3	5k	Adj.	1.97	N
To sneeze	2	2	8k	V	4.03	N
To snore	3	3	7k	V	4.39	N
Tenant	4	4	4k	N	4.55	N
Tonsils	6	6	12k	N	4.72	N
To totter	4	4	10k	V	–	N
To tuck in	2	2	4k	V	3.86	N
Upper berth	4	4	7k	N	3.37	N
To vacuum	2	2	5k	V	4.22	N

Notes:
*COCA: Corpus of Contemporary American English (Davies 2009).
**Concreteness ratings taken from Brysbaert, Warriner, and Kuperman (2014).